Data Analysis on Marine Engine Operating Regions in relation to Ship Navigation.

Lokukaluge P. Perera[1] and Brage Mo

Norwegian Marine Technology Research Institute (MARINTEK), Trondheim, Norway.

ABSTRACT

Data analysis techniques to understand marine engine operational regions as a part of the ship energy efficiency management plan (SEEMP) are proposed in this study. The SEEMP enforces to improve ship energy efficiency under various emission control measures by collecting and analyzing vessel performance and navigation data. The required data analysis techniques to analyze such data sets are presented under the engine-propeller combinator diagram (i.e. one propeller shaft with its own direct drive main engine). These techniques consist of implementing Gaussian Mixture Models (GMMs) with an Expectation Maximization (EM) algorithm to classify and Principal Component Analysis (PCA) to analyze frequent operating regions of a marine engine in a selected vessel. Three marine engine operating regions are noted under the combinator diagram and GMMs capture the shape, orientation and boundaries of those operating regions. Then, PCA is used to understand the structure of each GMM with respect to ship performance and navigation parameters. Hence, this approach can be used in the SEEMP to monitor ship navigation with respect to marine engine operating regions.

Keywords: SEEMP; EEOI; Ship Energy Efficiency; Gaussian Mixture Models; Expectation Maximization Algorithm, Principal Component Analysis.

1. Introduction

[1] Corresponding Author e-mail: prasad.perera@marintek.sintef.no; Tel: +47 464 15 000

1.1 Ship Energy Efficiency

Modern vessels are equipped with various onboard sensors and data acquisition (DAQ) systems to collect ship performance and navigation parameters. Such vessel related parameters are collected as large scale data sets that should be analyzed to evaluate vessel performance levels under various weather conditions. Ship speed, power and fuel consumption requirements are often studied under such vessel performance levels. Hence, marine engines in vessels can play an important role in such performance evaluation processes. However, various data analysis techniques should be developed to understand such vessel performance levels under ship operation data. Furthermore, these data analysis techniques and respective results can be a part of the ship energy efficiency management plan (SEEMP) and that help to draw conclusions on vessel performance levels.

It is believed that the SEEMP (IMO, 2009a and 2012) can play an important role in commercial ship operations in the future years as a mandatory mechanism in shipping. That enforces to improve vessel operational conditions and implement technology advancements for more energy efficient shipping fleets. The energy efficiency operational indicator (EEOI) can be used as a benchmark level for the SEEMP. Hence, the proposed data analysis techniques under vessel performance and navigation data facilitate for achieving the respective ship energy efficiency objectives that are assigned in the SEEMP (IMO, 2009b and 2012).

The SEEMP consists of four phases: planning, implementation, monitoring, self-evaluation and improvements. The first phase relates to vessel and shipping company specific measures, where goal setting type initiatives to improve ship energy efficiency should initiate with human resource development strategies of the shipping company. At the implementation phase, the same should be established with several implementation procedures by documenting the implementation progress and developing the required condition monitoring

(CM) facilities (Perera (2016a and 2016b) and Perera and Mo (2016a)). At the monitoring phase, ship performance and navigation data should collect and analyze to observe vessel energy efficiency under the implementation procedures. One should note that the proposed data analysis techniques can be used in this phase, extensively. At the self-evaluation phase, various voluntary reporting and review processes should be conducted and lessons learned should also be documented for further improvements. To achieve SEEMP objectives, the crew should have proper knowledge and training on energy management approaches in their vessel. These energy management approaches should be facilitated by onboard ship performance and navigation data and the respective data analysis techniques during the monitoring phase of the SEEMP. The final results of such data analysis techniques can increase the understanding of efficient operational conditions of the vessel.

1.2 The Recent Studies

This study proposes several data analysis techniques to understand operational regions of marine engines in relation to ship navigation as part of the SEEMP. A limited number of engine related ship navigation studies are presented in the recent literature (Armstrong and Banks, 2015) and summarized in this section. A comparison among four different procedures to optimize a combined Diesel and organic Rankine cycle system (i.e. inclusion of engine control variables) with the ship operational profile of a chemical tanker is presented by Baldi *et al.*, (2015). A coupled dynamic-thermodynamic engine simulation in time-domain is developed to evaluate engine performance across full range of operational conditions of a vessel by Murphy *et al.*, (2015). Furthermore, the respective engine performance curves for such marine engines under computer simulations are presented by Morsy *et al.*, (2011).

However, these studies are limited to various empirical models and such models often fail to accommodate large scale data sets of ship performance and navigation parameters. Therefore, this study proposes appropriate data analysis techniques to overcome such

situations. The proposed approach consists of understanding operational regions of a marine engine under the respective engine-propeller combinator diagram by considering ship performance and navigation data. Hence, the respective results can be used in the ship operation phase to identify optimal engine-propulsion operating conditions as a part of the SEEMP (Perera *et al.*, 2015a and 2015b).

Appropriate engine-propulsion configurations (i.e. optimal conditions) reduce respective power/fuel consumption and exhaust emissions in ships, significantly (Trodden *et al.*, 2015). That can be identified by observing vessel operations under engine-propeller combinator diagrams as further discussed in this study. The combinator diagram (i.e. the relationship between main engine (ME) power and shaft speed) facilitates as the basis to identify such appropriate engine-propulsion configurations under the respective data analysis techniques. Frequent operating regions of the engine-propulsion combinator diagram with the respective ship performance levels (i.e. speed and power conditions) can be identified under the same. The optimal engine operating regions can be extracted from such frequent operating regions under the same and that knowledge can be used to improve ship performance levels. Such optimal engine operating regions can be identified by observing data structures of each frequent operating region of the engine propeller combinator diagram. Hence, a ship performance and navigation data set of a selected vessel (i.e. one propeller shaft with its own direct drive main engine) is considering to implement the respective data analysis techniques as the main contribution of this study.

1.3 Engine Propeller Combination

Optimal engine-propulsion configurations are selected at the ship design phase with respect to vessel operational and navigation requirements. However, such configurations may degrade in ship navigation situations due to various environmental factors. The respective engine-propulsion interactions should be monitored in ship navigation and that information should be

analyzed to evaluate vessel performance. These engine-propulsion interactions are often studied under the combinator diagram as mentioned previously. Therefore, a general overview of the engine-propulsion combinator diagram is presented in Figure 1 consisting of a relationship between marine engine power (in a logarithmic scale) and shaft/propeller speeds. A direct drive situation, a marine engine (ME) is connected with a direct shaft to drive the propeller (i.e. fixed-pitch-propeller), is considered in this situation. In general, various ship navigation situations can be studied under such engine-propeller combinator diagrams (MAN Diesel & Turbo, 2011). Therefore, modern integrated bridge systems are often equipped with such onboard combinator diagrams to evaluate vessel performance levels in various environmental conditions. The respective engine fuel consumption (i.e. specific fuel consumption (SFC)) can also be incorporated into such combinator diagrams to evaluate real-time emissions in vessels.

The respective features of an engine-propeller combinator (see Figure 1) are further explained in this section. As presented in the figure, the maximum and minimum constant speed lines of (A1) and (A2) and mean effective pressure (MEP) lines of (A3) and (A4) limit the initial engine operation region. One should note that the engine idling/clutch-in speed range (A5) is close to (A2). The intersection between (A1) and (A3) represents the nominal maximum continuous rating (MCR) point (A5) for the respective engine. Further limitations of the engine operation region are the respective 110% engine power limit, sea trial engine speed limit (B1) and engine overload limit (B2). However, vessels operate beyond these engine limits with additional engine power capabilities under some sea trial situations. The engine operating region is further limited by the maximum acceptable engine speed limit (C1), maximum engine power limit (C2) (that can be 100% of respective engine power), 100% mean effective pressure limit (C3) and maximum torque-speed limit (C4).

The specified MCR point (P1) of the engine is located at the intersection of the 100 % engine power and speed limits. This limited engine operation region is divided into several circular regions (D1) of specific fuel consumption (SFC) levels. The smallest SFC region (i.e. small radius) represents the optimal fuel consumption rate (i.e. minimum SFC) of the engine. Therefore, the selection of propeller operating points in vessels should close to this optimal SFC region to reduce the respective fuel consumption for the same ship speed. In general, propeller configurations in vessels can be divided into two categories of: fixed-pitch-propellers (FPP) and controllable-pitch-propeller (CPP). A FPP design is for a specific operation speed with its optimal pitch condition. The CPP on the other hand can have additional flexibilities by changing its pitch condition with respect to engine speeds.

A light running FPP curve (E1) illustrates vessel operations under clean hull and propeller conditions in calm waters. The propeller design point (P2) and alternative propeller design point (P3) are also located in the same line. One should note that (P2) can move towards (P3) under the respective sea margin. i.e. the sea margin (F1) (i.e. up to 20% of respective engine power). However, the same curve can move towards the heavy FPP curve (E2) due to fouled hull and propeller conditions and/or rough weather, where (P2) (i.e. light running engine) can move towards (P4) (i.e. heavy running engine). A considerable ship speed reduction under engine heavy running conditions can be observed and a possible increase in the fuel consumption. Such heavy running situations can be avoided by appropriate dry-docking (i.e. cleaning hull and propellers) and voyage planning (i.e. weather routing and speed optimization) approaches (Fang and Lin, 2015). Similarly, (P4) can move towards (P5) under the sea margin (F1) under heavy running conditions of the engine. Furthermore, (P5) can move towards the specified MCR point for propulsion (P6) under the engine margin (F2).

If the marine engine has a shaft generator, then the engine should be able to supply an additional amount of power (F3) (i.e. (P6) should move towards (P1)) to the vessel. (P1) may

intercept the ultimate heavy FPP curve in (E3) as it relates to the ship specifications (MAN Diesel & Turbo, 2011). If the marine engine has a shaft motor, then an additional amount of power can be injected (i.e. (P6) should move opposite to (P1)) into the propulsion system by auxiliary engines. However, (P1) and (P6) can also coincide in the combinatory diagram in some vessels without PTO/PTI facilities. The recommended engine operating point with the lowest operation SFC value (i.e. the optimal fuel consumption) is (P7). Therefore, each engine operating point should move towards (P7) to improve energy efficiency in the vessel. However, the engine operating point may move towards (P8), the continuous service rating of the engine, as per the vessel operational requirements.

One should note that (P1), (P7) and (P8) may locate on the engine service curve, (G1), as presented in the figure. The FPP operating points can vary along approximate straight lines with respect to engine operating conditions. However, the CPP operating points can have additional flexibilities to move around the engine operating region, where the propeller pitch can vary with respect to required vessel and engine speeds. This flexibility in CPPs can be used to operate marine engines around their optimal operating regions, where the respective SFC can be reduced. However, both CPPs and FPPs may have additional operating limitations due to inception of the suction-side (H1) and pressure-side (H2) cavitation conditions and that possibly increases vessel fuel consumption. Furthermore, the same figure presents constant ship speed lines (I4) with respect to engine propeller operating conditions.

Ship performance and navigation data (i.e. in a selected vessel) on a similar engine propeller combinator diagram are considered to develop the proposed data analysis techniques. That consist of implementing Gaussian Mixture Models (GMMs) with an Expectation-Maximization (EM) algorithm on the combinator diagram to classify and Principal Component Analysis (PCA) to analyze frequent operating regions of the marine

engine. These mathematical techniques (i.e. GMMs, EM algorithm, and PCA) are further described in the following section.

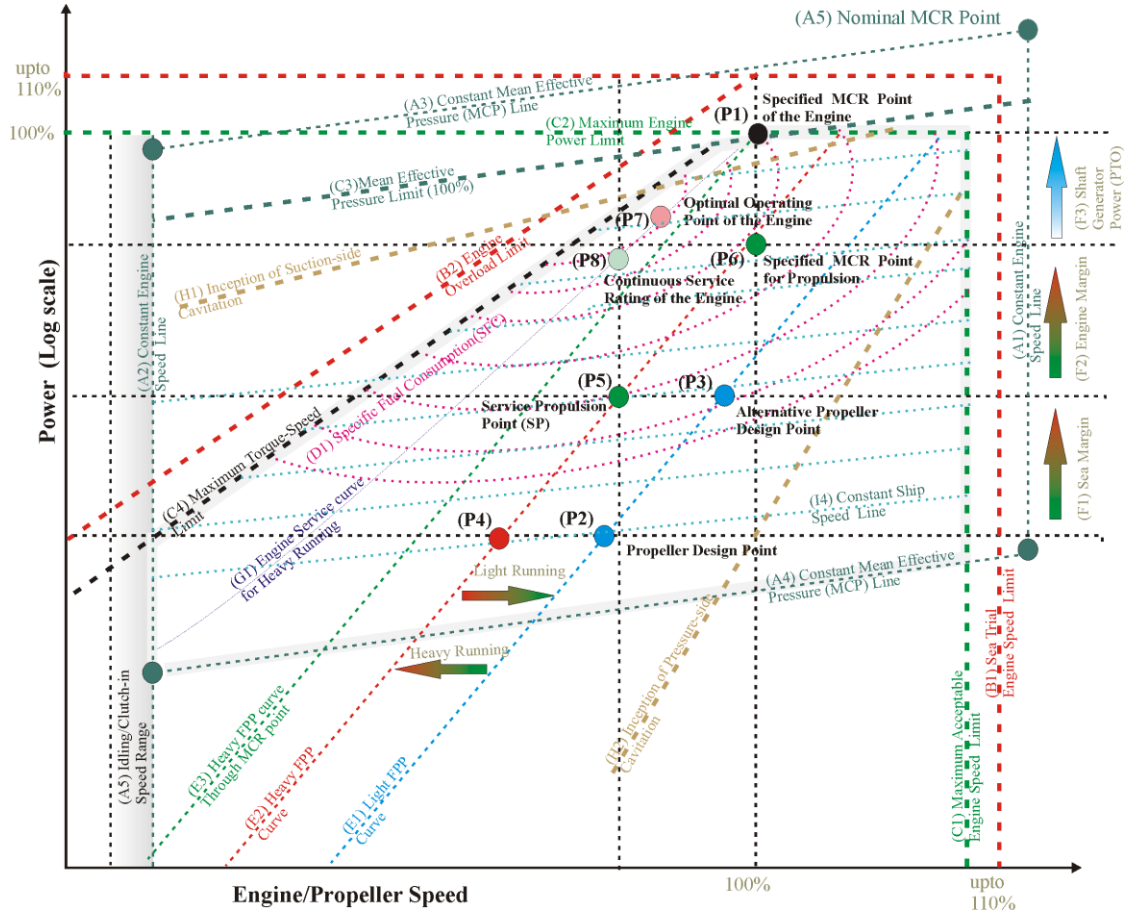

Fig. 1. Engine-propulsion combinator diagram

2.1 Gaussian Mixture Models

This section proposes to use GMMs (Sun *et al.*, 2006) to identify the most frequent operating regions of the engine-propeller combinator diagram at the first step. This method is a data clustering algorithm (clustering ship performance and navigation data), where the respective parameters of the GMMs are calculated by an iterative process (i.e. EM algorithm). Each frequent operating region of the engine-propeller combinator diagram is represented by a GMM and the respective parameters are calculated by the EM algorithm. Hence, each GMM of the combinator diagram has its own mean and covariance values that are approximated at the initial step. Each GMM (i.e. each data set) calculates its respective mean

and covariance values under the EM algorithm during the clustering process. However, several independent multivariate Gaussian distributions (i.e. GMM) in the engine-propeller combinatory diagram with respect to the engine operating points (i.e. engine modes) can be identified during this process.

2.2 Expectation-Maximization Algorithm

The data point assignment for each GMM is done by the EM algorithm (Moon, 1996), which consists of two iterative levels: expectation and maximization. The EM algorithm is used as an effective iterative procedure for maximum likelihood estimation (MLE) in this study. Therefore, it calculates the respective model parameters of the GMMs, as mentioned before. In the expectation step, the probability that each data point belongs to the respective data cluster (i.e. GMM) evaluates. In the maximization step, each data point accommodates in the data cluster with respect to the highest probability and updates the mean and covariance values of the respective GMM. This method assigns each data point exactly to one operational region (i.e. the respective GMM) of the engine-propeller combinatory diagram. Therefore, the boundaries of the most frequent operating region of the engine-propeller combinator diagram can be determined. The E-step is initiated by considering a multivariate GMM and denoted as (Ng, 2015):

$$p_j(x;\mu_j,\Sigma_j) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma_j\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)} \tag{1}$$

where x is the input data set and $p_j(x;\mu_j,\Sigma_j)$ is the PDF of the multivariate Gaussian distribution with, μ_j and Σ_j , the mean and covariance values of the of the j-th data cluster, respectively. The probability of i-th data point belongs to j-th cluster can be written as:

$$w_j^{(i)} = p\left(z^{(i)} = j \mid x^{(i)};\phi,\mu,\Sigma\right) \tag{2}$$

One should note that (2) calculates the "soft guess value" for the parameter, $z^{(i)}$. Considering the Bayes rule and (1), the posterior probability of the parameter, $z^{(i)}$, given the parameter, $x^{(i)}$, can be written as:

$$p\left(z^{(i)}=j \mid x^{(i)};\phi,\mu,\Sigma\right)=\frac{p\left(x^{(i)} \mid z^{(i)}=j;\mu,\Sigma\right)p\left(z^{(i)}=j;\phi\right)}{\sum_{l=1}^{k} p\left(x^{(i)} \mid z^{(i)}=l;\mu,\Sigma\right)p\left(z^{(i)}=l;\phi\right)} \tag{3}$$

where $p\left(z^{(i)}=j;\phi\right)$ is the prior probability of the j-th data cluster and k is the number of data clusters. The equal prior probability of each data cluster is assumed at the initial step of the EM algorithm. One should note that (3) represents a multivariate Gaussian distribution with μ_j and Σ_j are the mean and covariance values, respectively. The respective M-step can be written as:

$$\begin{aligned}
&\phi_j=\frac{1}{m}\sum_{i=1}^{m} w_j^{(i)} \\
&\mu_j=\sum_{i=1}^{m} w_j^{(i)} x^{(i)} \Big/ \sum_{i=1}^{m} w_j^{(i)} \\
&\Sigma_j=\sum_{i=1}^{m} w_j^{(i)}\left(x^{(i)}-\mu_j\right)\left(x^{(i)}-\mu_j\right)^T \Big/ \sum_{i=1}^{m} w_j^{(i)}
\end{aligned} \tag{4}$$

This step is used to update the respective data cluster (i.e. the GMM) by calculating the new mean and covariance values with respect to each data point. This iterative process should stop either at the end of the training data set or when approximately stable prior and posterior to mean and covariance values. However, the EM algorithm may converge to a local minima or saddle point in some situations during its iterative process. Hence, the initial mean and covariance values should be selected, appropriately.

2.3 Principal Component Analysis

This section proposes to use PCA to analyze frequent operating regions of the engine-propeller combinator diagram that are identified in the previous section. As the next step, the structure of each data cluster (i.e. GMM) under the engine propeller combinator diagram is identified by PCA (Jackson, 1980). Hence, a classified data set of ship performance and

navigation parameters is considered, where $x_1(t), x_2(t), \ldots, x_m(t)$ denote the respective parameters. The sample mean, $\bar{x}$, and variance, S_x, of the same can be written as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \tag{5}$$

That is into a new data set transformed by considering the following steps under PCA:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = u^T \bar{x}$$

$$S_y = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T = u^T S_x u \tag{6}$$

where $\bar{y}$ is the mean of the new data set, and S_y is the respective variance of the transformed data set and u is a unit variance vector that also satisfies:

$$u^T u = I \tag{7}$$

PCA maximizes the value of each variance direction (i.e. principal component direction) of the new data set. Hence, the trace of S_y should be maximized as

$$\text{Max. } trace(S_y) = \text{Max. } trace(u^T S_x u) \tag{8}$$

The Lagrange multiplier that satisfy (8) can be written as:

$$L = trace(u^T S_x u) = \sum_{i=1}^{n}\left[u_i^T S_x u_i + \lambda(1 - u_i^T u_i)\right] \tag{9}$$

The derivatives of the Lagrange multiplier in (9) can be resulted in:

$$S_x u_i = \lambda_i u_i$$

$$u_i^T u_i = 1 \tag{10}$$

where λ_i is the eigenvalues and u_i is the respective eigenvectors of S_x. Hence, the respective eigenvalues and eigenvectors help to identify the hidden structure of the respective data sets that are clustered around the engine propeller combinator diagram.

3 Data Analyses

3.1 Engine Centered Approach

This section consists of implementing Gaussian Mixture Models (GMMs) with Expectation Maximization (EM) algorithm to classify and Principal Component Analysis (PCA) to analyze frequent operating regions of marine engines. That is done by considering a data set of ship performance and navigation parameters with respect to the engine propeller combinator diagram. To improve ship energy efficiency, such approach should also be done by monitoring ship performance and navigation parameters in real-time with respect to marine engine operational regions, where the respective SFC should be minimized. One should note that marine engines operate around selected RPM values (i.e. engine modes) that relate to the required ship speeds. These operating values are also related to engine loading conditions, where each marine engine has several mean RPM values (i.e. Engine modes) that implement to achieve required ship speeds as mentioned before. These engine related operating conditions can be captured by ship performance and navigation data as discussed. It has shown that the respective ship performance and navigation parameters should be classified along the engine operating regions to identify such mean RPM values. Hence, the proposed data analysis techniques support of such data classification approach to visualize marine engine operating regions.

Various data analysis techniques often use as decision support tools in the maritime and offshore industries (Perera *et al.,* 2012). These decision support tools are based on large scale data sets and these data sets may introduce additional challenges in the respective data analysis techniques. In general, these large data sets of ship performance and navigation parameters may consist of some erroneous data intervals due to sensor and DAQ fault/noise conditions. However, such erroneous conditions can also be identified, effectively by the proposed data classification (i.e. data clustering) and structure identification approaches because the respective errors are more visible in small data sets, effectively. Hence, these

erroneous conditions in ship performance and navigation parameters should be removed to create cleaner data sets (Perera and Mo, 2016b). The cleaned data sets (i.e. without erroneous data conditions) can be used to monitor ship performance under the SEEMP.

3.2 Ship Performance & Monitoring Data

Ship performance and navigation data of a selected vessel are considered in this section and that is collected in one year period, approximately. The vessel consists of following particulars: ship type: bulk carrier, ship length: 225 (m), beam: 32.29 (m), gross tonnage: 38.889 (tons), deadweight at max draft: 72.562 (tons). The vessel is powered by 2 stroke marine engine with maximum continuous rating (MCR) of 7564 (kW) at the shaft rotational speed of 105 (rpm). The vessel has a fixed pitch propeller with diameter 6.20 (m) with 4 blades.

Initially, a three parameter data set that relates to the engine of the vessel is selected and the respective parameters are categorized as: average (avg.) draft (m), speed through water (STW) (Knots), main engine (ME) power (kW), shaft speed (rpm), main engine (ME) fuel consumption (cons.) (Tons/day), speed over ground (SOG) (Knots), trim (m), relative (rel.) wind speed (m/s) and direction (deg) and auxiliary (aux.) fuel consumption (cons.) (Tons/day). Further details on the respective sensors and DAQ system used to collect these ship performance and navigation parameters are presented in Perera *et al.,* (2015a). Furthermore, the following parameter ranges are considered: ME power from 3000 (kW) to 8000 (kW) and shaft speeds from 80 (RPM) to 120 (RPM). In general, these values represent frequent engine propeller operating regions in the combinator diagram. The data points near low speed-power values that represent slow moving situations of the vessel are ignored from the data analysis. Therefore, considerable sensor and DAQ noise data (i.e. erroneous data intervals) are removed by that step.

An initial statistical analysis of parameters that relate to general engine-propeller operating situations is presented in Figure 2. The top, middle and bottom plots of the figure represent various histograms of the parameters of shaft speed, ME power and fuel consumption, respectively. Three frequent operating regions are observed in this statistical analysis and relate to three engine operating modes (i.e. data clusters 1, 2 and 3). Hence, the results show that the engine operates around three RPM regions (i.e. approximately Gaussian distributions) and creates the respective power and fuel consumption regions. The mean and covariance values for these approximately Gaussian distributions are calculated by considering the same figure.

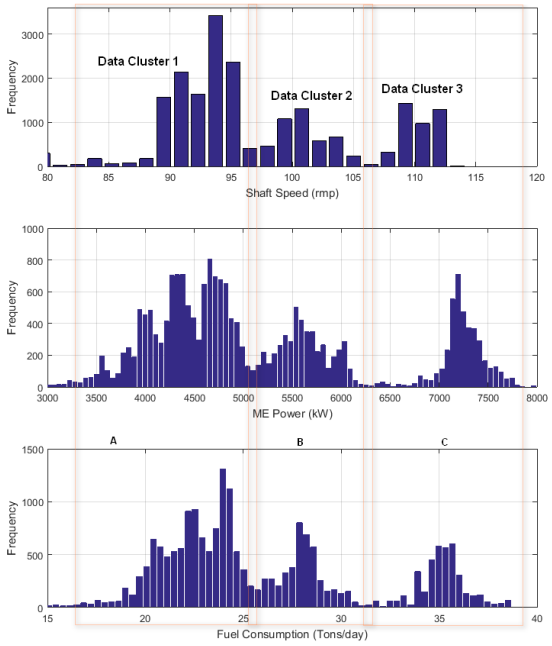

Fig. 2. Statistical distributions of engine parameters

As the next step, the two parameters (i.e. ME power and shaft speed) are combined to develop an engine-propeller combinator diagram and the results are presented in Figure 3. The bottom plot of the same figure represents the histogram of combined ME power and shaft speed values. One should note that the same engine operating regions (see Figure 2) are also noted in this combinator diagram (i.e. data clusters 1, 2 and 3). The top-left plot of the same figure represents the contours of the bottom plot. The top-right plot of the same figure

represents the same contours with the respective fuel computation values. In general, higher ME power-shaft speed regions consist of higher fuel consumption values and vice versa. However, higher fuel consumption values in lower ME power-shaft speed regions can also be noted in some situations due to rough ocean conditions.

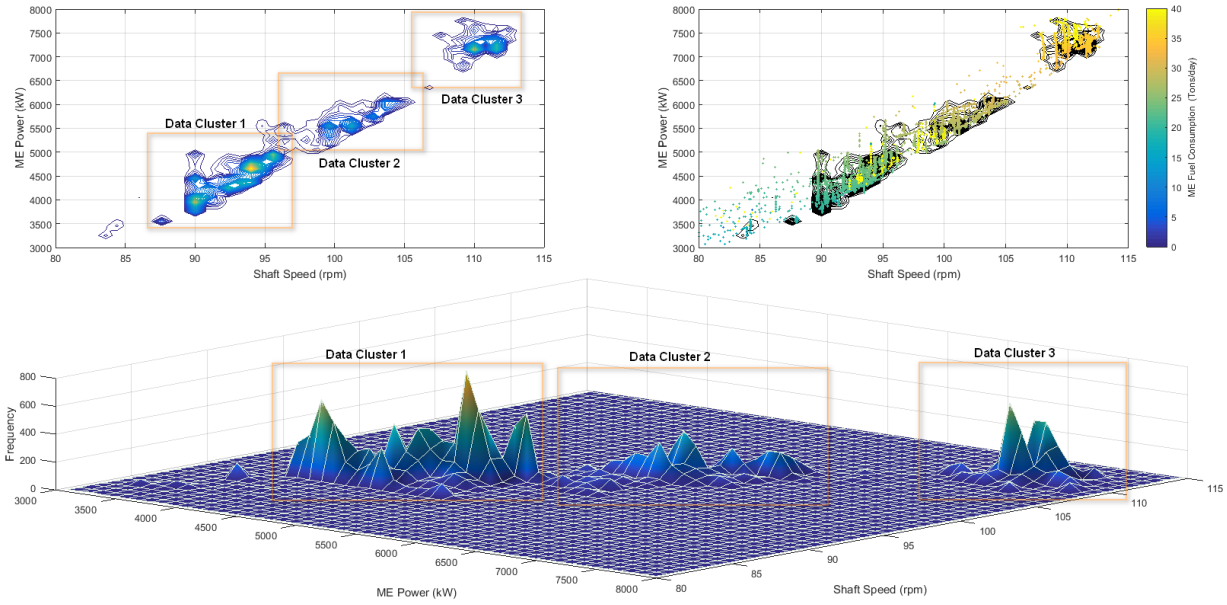

Fig. 3. Engine operation regions: ME Power vs. Shaft Speed

The same contour plot with respect to relative (Rel.) wind speed is presented in the top left plot of Figure 4. It is assumed that the relative wind conditions relate to encountered sea states in the respective ship route. The results show that shaft speeds of the engine have reduced significantly under the same power level because of high wind speeds (i.e. higher engine loading conditions). The propeller is rotating at relatively slow speeds in such situations due to engine loading conditions, where ship speed is also degraded. The same contour plot with respect to speed through water (STW) of the vessel is presented in the top right plot of Figure 4. These two plots show that engine power increments improve STW and rough weather conditions (i.e. high wind speeds) degrade STW, significantly (Lu *et al.*, 2015). The same contour plots with respect to vessel trim and average (avg.) draft values are presented in the bottom left and right plots of Figure 4. One should note that these avg. draft values relate to

loading conditions of the vessel, therefore appropriate trim values are selected to reduce ship resistance during these ship navigation situations.

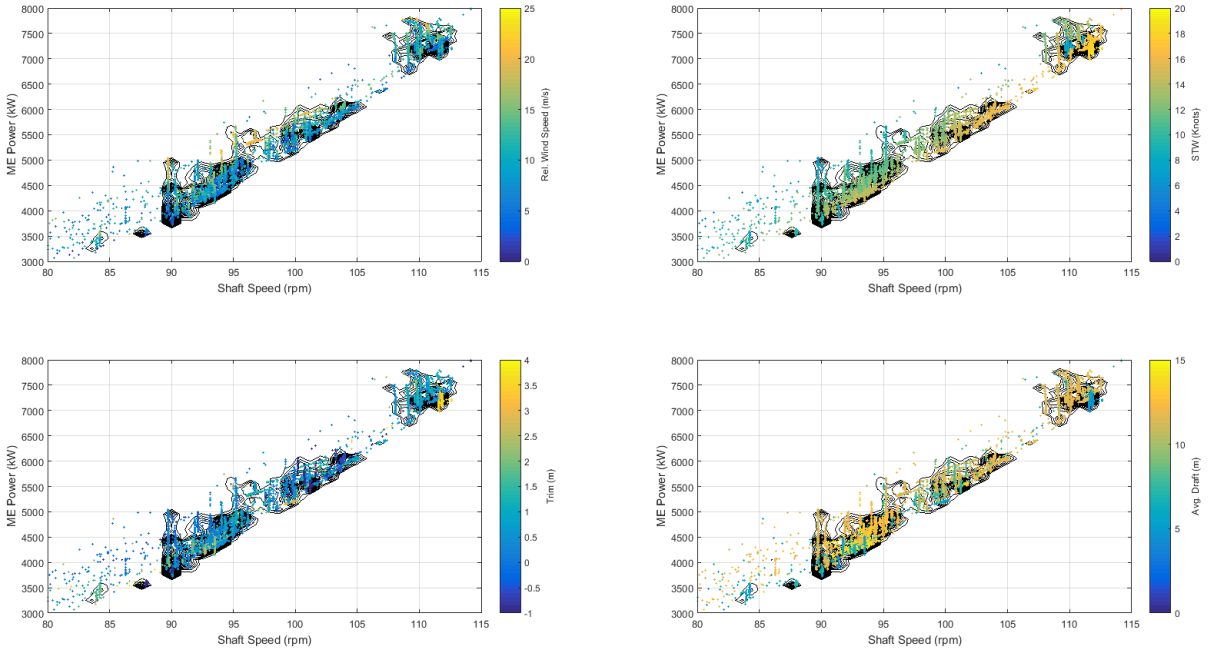

Fig. 4. Engine operation region vs. rel. wind speed, avg. draft, trim and log speed

Another view of the engine propeller combinator diagram is presented in Figure 5. The respective operating patterns of the propeller are identified as straight lines and those are also marked as gray lines in the same. These straight lines are observed during a real-time simulation of the same data set. The ME power axis (i.e. y-axis) is presented in a log scale to improve the visibility of the respective engine-propeller data. The vertical gray lines represent various engine power values with constant shaft speed situations. However, the inclined gray lines represent continuous engine operating situations under varying sea conditions. One should note that these inclined lines are approximately similar to the fixed-pitch-propeller (FPP) lines in Figure 1 (i.e. E1, E2, and E3). These results confirm that the FPP operating points can vary along approximately straight lines with respect to engine running conditions. The same combinator diagram with the vessel STW values is presented in Figure 6. The results show that vessel STW degrades along the FPP lines (i.e. for the same ME power, the shaft speed degrades) from right to left. Therefore, an overview of various interactions among

marine engine, propeller, ship resistance, and environmental conditions can be observed in such combinator diagrams and that is a good tool to evaluate vessel performance. As the next step of this study, three operating regions (i.e. engine modes) noted in Figures 2 and 3 are identified as data clusters, where GMMs with the EM algorithm are implemented.

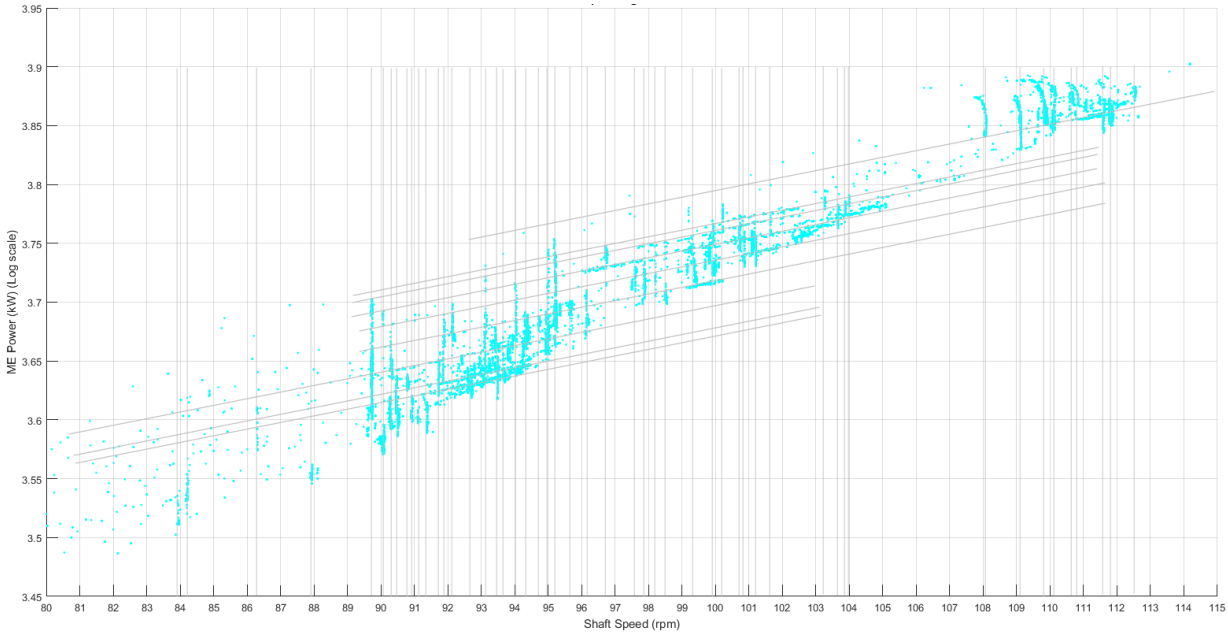

Fig. 5. Engine propeller combinator diagram

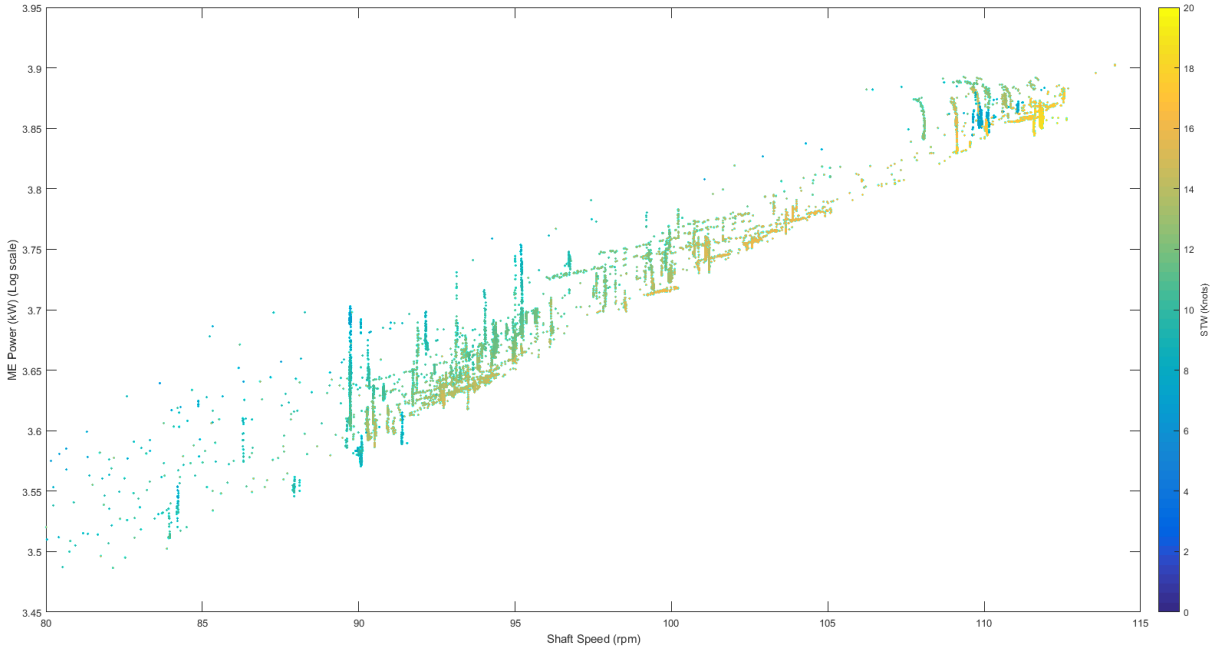

Fig. 6. Engine propeller combinator diagram with STW

3.3 GMM with EM algorithm

The initial mean and covariance values for GMMs are approximated from the statistical distributions in Figure 2. The estimated mean and variance values for each GMM are approximated as:

$$\mu_1 = [92.5(rpm) \quad \log(4500)(kW - \log scale)]$$

$$\mu_2 = [101(rpm) \quad \log(5500)(kW - \log scale)]$$

$$\mu_3 = [111(rpm) \quad \log(7300)(kW - \log scale)]$$

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 4.6106(rpm^2) & 0.0551(rpm.kW - \log scale) \\ 0.0551(rpm.kW - \log scale) & 0.0007(kW - \log scale)^2 \end{bmatrix} \quad (5)$$

These values are introduced as initial GMMs and presented in left plot of Figure 7 (i.e. the combinator diagram). These GMMs are selected near the local maxima points to improve the EM algorithm performance and represented as multivariate Gaussian distributions with the respective mean and covariance values in (5). The respective contours of multivariate Gaussian distributions (i.e. GMMs) are denoted by ellipse in this figure. Then, the EM algorithm is executed to update these GMMs with the respective data points and the results are presented in the right plot of Figure 7.

The results in the figure show that the GMMs converge to appropriate mean and covariance values under the EM algorithm. Hence, these three regions are identified as the most frequent operating regions of the engine-propeller combinator diagram by considering the respective ship performance and navigation data. The same regions are denoted as data clusters 1, 2, and 3 and that also relate to Figures 2 and 3. However, an overlay situation within two GMMs is also observed in these results. The calculated mean and covariance values of the respective GMMs are calculated as:

$$\mu_1 = [92.8770(rpm) \quad 3.6515(kW - \log scale)]$$

$$\mu_2 = [101.5521(rpm) \quad 3.7549(kW - \log scale)]$$

$$\mu_3 = [110.4419(rpm) \quad 3.8609(kW - \log scale)]$$

$$\Sigma_1 = \begin{bmatrix} 8.8222(rpm^2) & 0.1159(rpm.kW - \log scale) \\ 0.1159(rpm.kW - \log scale) & 0.0021(kW - \log scale)^2 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 3.6268(rpm^2) & 0.0340(rpm.kW - \log scale) \\ 0.0340(rpm.kW - \log scale) & 0.0004(kW - \log scale)^2 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 1.4695(rpm^2) & 0.0030(rpm.kW - \log scale) \\ 0.0030(rpm.kW - \log scale) & 0.0001(kW - \log scale)^2 \end{bmatrix} \quad (6)$$

One should note that slight variations among the initial and final means values and considerable variations among the initial and final covariance values are observed in these results. Hence, those GMMs (data cluster 1, 2, and 3) in the respective engine propeller combinator diagram represent frequent operating regions of the marine engine.

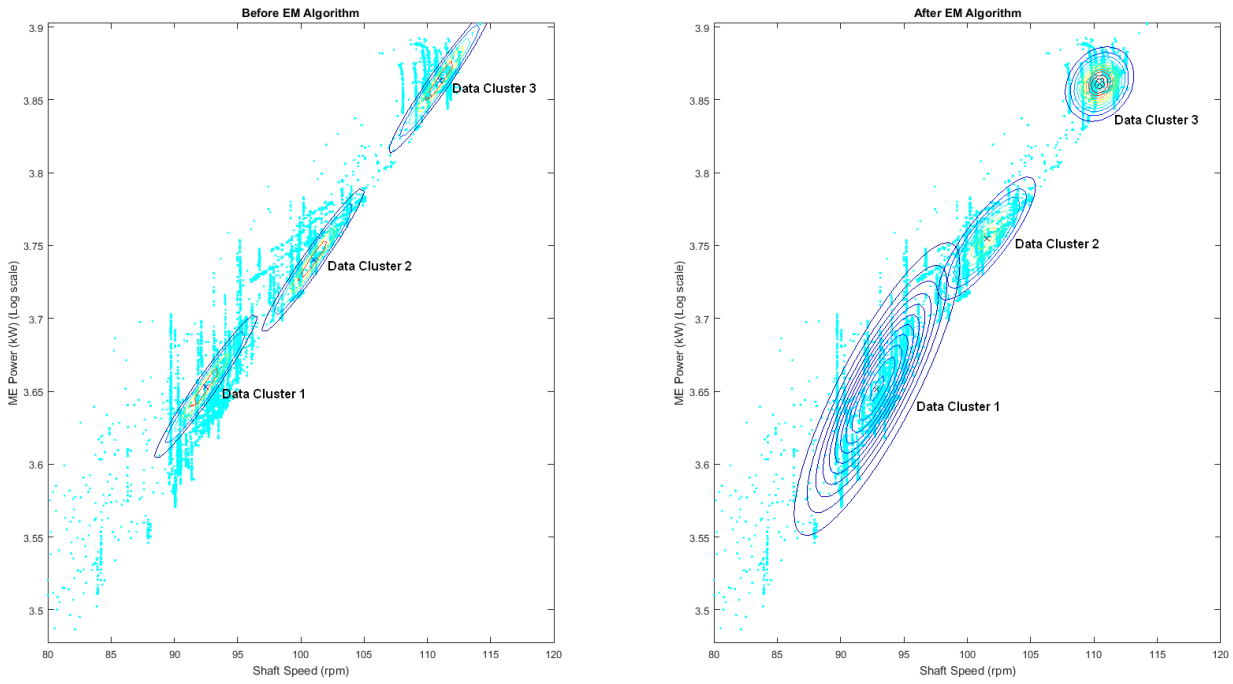

Fig. 7. Data Classification in the Engine propeller combinator diagram

One should note that these frequent operating regions have different variance values and that represent how such data sets are distributed. Furthermore, sensor and DAQ erroneous regions on the same data set can also be filtered this approach by introducing threshold values for the GMMs. Then the most important data points that relate to each GMM can be identified by these statistical distributions to reduce the respective data erroneous conditions. Hence, the calculated mean and covariance values of GMMs facilitate towards such data filters. It is also believed that these data regions can be used to derive localized ship performance and navigation models. In the next section, PCA is used to analyze a selected frequent operating

region (i.e. a GMM) of the marine engine, where linearized relationships among ship performance and navigation parameters are investigated.

3.4 PCA Analysis

A selected GMM (i.e. data cluster 3) in the engine propeller combinator diagram is considered for PCA in this section. The calculated PCs are presented in Figure 8, where the i-th PC is denoted as:

$$Z_i = \begin{bmatrix} z_{i,1} & z_{i,2} & z_{i,3} & z_{i,4} & z_{i,5} & z_{i,6} & z_{i,7} & z_{i,8} & z_{i,9} & z_{i,10} \end{bmatrix} \quad (11)$$

Therefore, where $z_{i,1}, z_{i,2}, z_{i,3}, .z_{i,10}$ represent the respective vector components of the i-th PC. One should note that the top and bottom PCs are Z_1 and Z_{10}, respectively. Hence, the respective vector components of each PC are further in the next step investigated by an appropriate data visualization approach and the results are presented in the same figure. This figure represents a 10 dimensional vector space, where the respective PCs (i.e. eigenvectors) are in a polar coordinates. Each PC is by a dotted circle presented, where the top PC has the largest radius. Each axis that is intersecting these circles represents a parameter from ship performance and navigation data. The respective vector components of each PC are by colored circles presented and the circle radius represents the significance of that component with respect to other components within the same PC. This figure also represents an overview of the correlations among the respective parameters of ship performance and navigation data. High positive (HP) correlations represented by yellow color large circles and high negative (HN) correlations represented by blue color large circles in Figure 8 (see the color bar).

The respective PCs with their vector components are in this section further discussed (see Figure 3). The 1st PC represents: when avg. draft (high) increases (HP), STW (medium) decreases, shaft speed (medium) decreases, SOG (high) decreases, and trim (medium) decreases. The 1st PC shows that ship resistance has increased due to the draft increments, where STW and SOG of the vessel are also decreased. The same conditions have decreased shaft speeds due to high engine loading conditions. Furthermore, the draft increments are compensated by trim adjustments of the vessel and that is by this PC also noted. In general, when vessel avg. draft increases, then trim increases, STW decreases, and SOG decreases.

The 2nd PC represents: when engine power and shaft speed (medium) increases, ME fuel consumption (high) increases, and aux. fuel consumption (high) increases. The 2nd PC shows a moderate increment in engine shaft speed increases engine power levels, moderately and fuel consumption in both main and auxiliary engines, significantly. This results show that shaft speed increments beyond the mean operating point in this engine operating region may not increase engine power, considerably but that may increase the respective fuel consumption, significantly. In general, when engine power and shaft speed increase, ME fuel consumption increases, and aux. fuel consumption increases.

The 3rd PC represents: when rel. wind speed (medium) decreases, then rel. wind angle (high) increases. Therefore, the 3rd PC shows that when the vessel increases its speed, then rel. wind speed increases and rel. wind angle decreases (i.e. the vessel encounters high head wind conditions with the speed increments). The 4th PC represents: when ME power(high) increases, then shaft speed (medium) increases. Therefore, the forth PC shows, the shaft speed increments increase engine power. The 5th PC represents: when vessel trim (medium) increases, then relative wind speed (high) and direction (high) increase. The 5th PC shows that the trim values are used under calm water conditions, where relative wind speed is slower and wind angle is higher. A large wind angle represents a situation, where the vessel is moving in moderate or slow speeds and not encounter any high head winds. The 6th PC represents: when STW (high) decreases, then relative wind angle (medium) decreases. The 6th PC shows, a positive correlation between STW and relative wind direction of the vessel and that relationship is similar to the previous PC. The 7th PC represents: when avg. draft (medium) increases, then SOG (medium) decrease, trim (high) increases, and rel. wind speed (medium) decreases. Similarly, the 7th PC shows that ship resistance has increased due to the draft increments, therefore SOG is also decreased. The same conditions have reduced rel. wind speed, as discussed previously. Furthermore, draft increments are compensated by trim variations under slow maneuvering conditions of the vessel in such situations.

The 8th PC represents: when avg. draft (high) increases, then shaft speed (medium) increases. The 8th PC shows that ship resistance has increased due to draft increments, therefore shaft speed is increased to compensate speed losses in the vessel. The 9th PC represents: when shaft speed (high)

decreases, then SOG (high) decreases. The 10th PC represents: when ME fuel consumption (high) decreases, then aux. fuel consumption (high) increases. The bottom PCs may not represent any useful information about the respective parameter relationships as mentioned before. Therefore, a proper interpretation for the bottom PCs should not be expected. Furthermore, that can accumulate data erroneous conditions of ship performance and navigation information, therefore such parameter relationships should be ignored. The low positive and negative correlations among the respective parameters are from the above discussion ignored. Those effects (i.e. low positive and negative correlations) should also be incorporated in the respective parameter relationship to see an overall picture of ship performance and navigation information. However, that can complicate the outcome of the respective PCs in some situations.

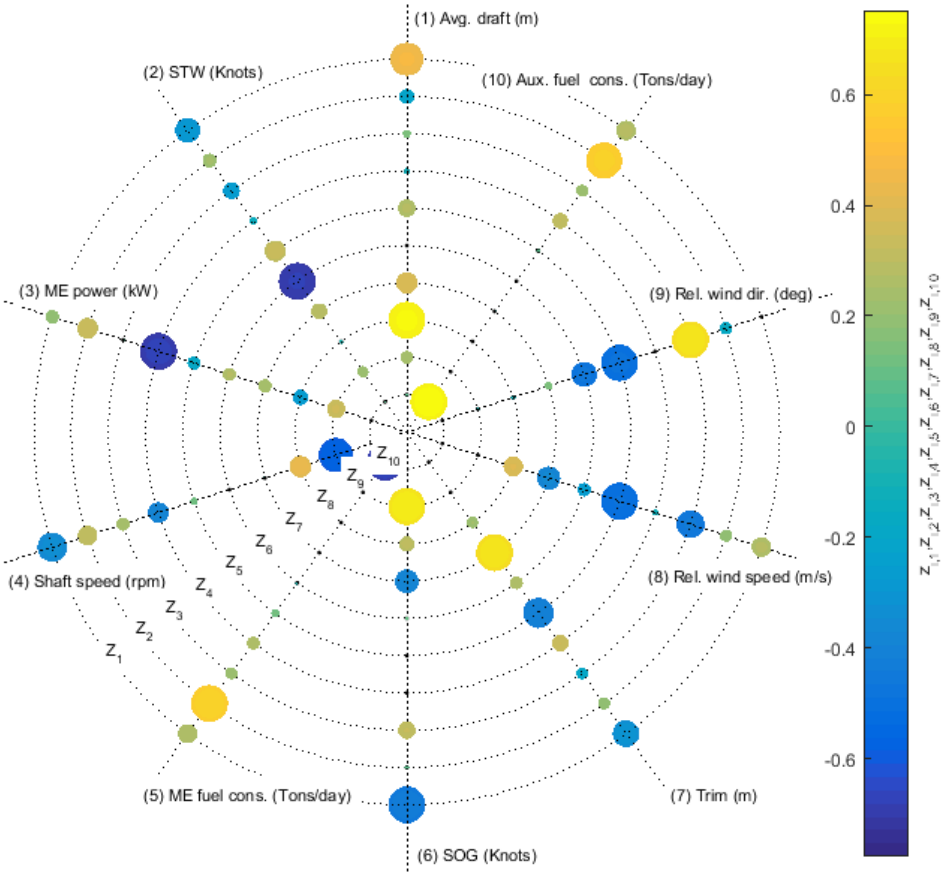

Fig. 8 PCs for Data Cluster 3

4 Conclusion and Future Work

This study consists of implementing GMMs with the EM algorithm to classify and PCA to analyze frequent operating regions of marine engines in relation to ship navigation. Three marine engine operating regions are noted under the engine propeller combinator diagram, where GMMs with the EM algorithm captures the shape, orientation and boundaries of those operating regions. Then, PCA is used to understand the structure of a selected GMM with respect to ship performance and navigation parameters. This approach is categorized as the development of data analysis techniques under the engine propeller combinator diagram to evaluate vessel performance in large scale data sets. These operating regions (i.e. engine modes) under the proposed data analysis techniques can be used to evaluate vessel performance under the SEEMP to overcome the current shipping industrial challenges in emission control based energy efficiency measures (Perera and Mo, 2016c, 2016d). Hence, the respective marine engine and propeller operating regions can be selected appropriately with respect to the PCs to reduce the fuel consumption of the vessel. Furthermore, advanced mathematical models of ship performance monitoring will be developed under such operating regions by considering the respective PCs of the marine engine as the future work of this study.

Acknowledgement

This work has been conducted under the project of "SFI Smart Maritime (237917/O30) - Norwegian Centre for improved energy-efficiency and reduced emissions from the maritime sector" that is partly funded by the Research Council of Norway. The authors would like to thank anonymous reviewers for their helpful and constructive comments that greatly contributed to improve the final version of the paper.

High resolution color images of this data analysis are presented in: http://bit.do/perera. An initial version of this paper is presented at the 35th International Conference on Ocean, Offshore and Arctic Engineering, 2016, Busan, Korea.

References

Armstrong, V.N. and Banks, C., (2015), Integrated approach to vessel energy efficiency, Ocean Engineering, Volume 110, Part B, pp. 39-48.

Baldi, F., Larsen, U., and Gabrielii, C., (2015), Comparison of different procedures for the optimisation of a combined Diesel engine and organic Rankine cycle system based on ship operational profile, Ocean Engineering, Volume 110, Part B, pp. 85-93.

Fang, M.-C., and Lin, Y.-H., (2015) "The optimization of ship weather-routing algorithm based on the composite influence of multi-dynamic elements (II): Optimized routings," Applied Ocean Research, vol. 50, pp. 130-140.

Jackson, J. E. (1980). "Principal components and factor analysis: part i-principal components." Journal of Quality Technology, vol. 12, no. 4, pp. 201–213.

IMO, (2009) "Resolution MEPC.1/Circ.683," Guidelines for the development of a ship energy efficiency management plan (SEEMP).

IMO, (2009) "Resolution MEPC.1/Cric.684," Guidelines for the voluntary use of the ship energy efficiency operational indicator (EEOI).

IMO, (2012) "Resolution MEPC.213(63)," 2012 Guidelines for the development of a ship energy efficiency management plan (SEEMP).

Lu, R., Turan, O., Boulougouris, E., Banks, C. and Incecik, A., (2015), A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping, Ocean Engineering, Volume 110, Part B, pp. 18-28.

MAN Diesel & Turbo, (2011) "Basic principles of ship propulsion," Copenhagen, Denmark.

Morsy, M., Gohary, E., and Abdou, K. M., (2011), Computer based selection and performance analysis of marine diesel engine, Alexandria Engineering Journal, vol. 50, no. 1, pp. 1-11.

Moon, T.K., (1996) "The expectation-maximization algorithm," in Signal Processing Magazine, IEEE , vol. 13, no. 6, pp. 47-60, Nov 1996.

Murphy, A.J., Norman, A.J., Pazouki, K., and Trodden, D.G. (2015), Thermodynamic simulation for the investigation of marine Diesel engines, Ocean Engineering, Volume 102, pp. 117-128.

Ng, A., (2015) "Mixtures of Gaussians and the EM algorithm," Lecture notes on Machine Learning, Stanford University, USA.

Perera, L.P., Rodrigues, J. M., Pascoal, R., and Guedes Soares, C., (2012) "Development of an onboard decision support system for ship navigation under rough weather conditions,", Sustainable Maritime Transportation and Exploitation of Sea Resources, E. Rizzuto & C. Guedes Soares (Eds.), Taylor & Francis Group, London, UK, pp. 837-844.

Perera, L.P. (2016a) "Statistical Filter based Sensor and DAQ Fault Detection for Onboard Ship Performance and Navigation Monitoring Systems," In Proceedings of the 8th IFAC Conference on Control Applications in Marine Systems (CAMS 2016), Trondheim, Norway.

Perera, L.P. (2016b) "Marine Engine Centered Localized Models for Sensor Fault Detection under Ship Performance Monitoring," In Proceedings of the 3rd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology, (AMEST'16), Biarritz, France.

Perera, L.P., Mo, B., Kristjansson, L.A., Jonvik, P.C., and Svardal, J.O. (2015a) "Evaluations on Ship Performance under Varying Operational Conditions," In Proceedings of the 34th International Conference on Ocean, Offshore and Arctic Engineering, Newfoundland, Canada,(OMAE2015-41793).

Perera, L.P., Mo, B., and Kristjansson, L. A., (2015b) "Identification of Optimal Trim Configurations to improve Energy Efficiency in Ships," In Proceedings of the 10th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC 2015), Copenhagen, Denmark, pp. 267-272.

Perera, L.P., and Mo, B., (2016a), "Data Compression of Ship Performance and Navigation Information under Deep Learning," In Proceedings of the 35th International Conference on Ocean, Offshore and Arctic Engineering, Busan, Korea, (OMAE2016-54093).

Perera, L.P. and Mo, B., (2016b) ”Ship Speed Power Performance under Relative Wind Profiles,” Maritime Engineering and Technology III , Guedes Soares & Santos (Eds.), vol. 1, Taylor & Francis Group, London, UK, ISBN 978-1-138-03000-8, pp. 133-141.

Perera, L.P., and Mo, B. (2016c) ”Machine Intelligence for Energy Efficient Ships: A Big Data Solution,” Maritime Engineering and Technology III , Guedes Soares & Santos (Eds.), vol. 1, Taylor & Francis Group, London, UK, ISBN 978-1-138-03000-8, pp. 143-150.

Perera, L.P., and Mo, B. (2016d) ”Marine Engine Centered Data Analytics for Ship Performance Monitoring,” Journal of Offshore Mechanics and Arctic Engineering-Transactions of The ASME, (DOI: 10.1115/1.4034923).

Sun, S., Zhang, C., and Yu, G., (2006),"A Bayesian network approach to traffic flow forecasting," in Intelligent Transportation Systems, IEEE Transactions on, vol.7, no.1, pp. 124-132.

Trodden, D.G., Murphy, A.J., Pazouki, K., and Sargeant, J., (2015), Fuel usage data analysis for efficient shipping operations, Ocean Engineering, Volume 110, Part B, Pages 75-84.